View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarWorks @ Georgia State University

Georgia State University
ScholarWorks @ Georgia State University

Philosophy Theses

Department of Philosophy

8-13-2019

# Butler and Kant on Human Nature and Morality

Botian Liu

Follow this and additional works at: https://scholarworks.gsu.edu/philosophy_theses

Recommended Citation

Liu, Botian, "Butler and Kant on Human Nature and Morality." Thesis, Georgia State University, 2019.
https://scholarworks.gsu.edu/philosophy_theses/254

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

BUTLER AND KANT ON HUMAN NATURE AND MORALITY

by

BOTIAN LIU

Under the Direction of Eric Wilson, PhD

ABSTRACT

Kant and Butler have a sharp methodological conflict in justifying moral obligations. While Kant argues that moral obligations can only be grounded in *a priori* justifications rather than in anything empirical, Joseph Butler grounds moral obligations in the empirical knowledge of human beings. Despite the apparent radical difference, I argue that Kant agrees with Butler that moral obligations must be grounded in the understanding of human beings. They, however, fundamentally disagree about human nature, which generates their methodological conflict in studying morality. For Kant, the essential attribute for human beings is autonomy, which presupposes independence from any particular experience. In contrast, Butler understands human nature as a system that includes different particular experience. Although there is no conclusive answer of the correct understanding of human nature, I suggest that Butler's account of moral obligations is a plausible one that can be considered as a counterexample to Kant's account.

INDEX WORDS: Butler, Kant, Morality, Human Nature

# BUTLER AND KANT ON HUMAN NATURE AND MORALITY

by

BOTIAN LIU

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2019

Copyright by
Botian Liu
2019

# BUTLER AND KANT ON HUMAN NATURE AND MORALITY

by

BOTIAN LIU

Committee Chair: Eric Wilson

Committee: Andrew I. Cohen

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2019

## DEDICATION

I would like to dedicate my thesis to my parents and my wife, who make doing philosophy possible for me.

## ACKNOWLEDGEMENT

I would like to acknowledge my thesis advisor, Dr. Eric Wilson, for all his inspiration and help on this thesis. His instruction, encouragement, and support allowed me to successfully complete this project. I would also like to give acknowledgment to all those who looked through drafts of this paper and gave me invaluable feedback on it.

# TABLE OF CONTENTS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| CrPracR | *Critique of Practical Reason*, ed. Mary Gregor. (Cambridge University Press, 2015). |
| G | *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor and Jens Timmermann. (Cambridge University Press, 2012). |
| LE | *Lectures on ethics*, ed. Peter Heath and J.B. Schneewind. Translated by Peter Heath. (Cambridge University Press, 1997). |
| MM Kant | *The Metaphysics of Morals*, ed. Lara Denis. Translated by Mary Gregor. (Cambridge University Press, 2017). |

# 1 INTRODUCTION

> There are two ways in which the subject of morals may be treated. One begins from inquiring into the abstract relations of things: the other, from a matter of fact, namely, what the particular nature of man is, its several parts, their economy or constitution; from whence it proceeds to determine what course of life it is, which correspondent to this whole nature…The following discourses proceed chiefly in this later method. The three first wholly. They were intended to explain what is meant by nature of man, when it is said that virtue consists in following, and vice in deviating from it; and by explaining to that the assertion is true.
>
> ---Joseph Butler[1]

> The ground of the obligation here must not be sought in the nature of the human being, or in the circumstances of the world in which he is places, but a priori solely in concepts of pure reason, and that any other prescription that is in some certain respect universal, in so far as it relies in the least part on empirical grounds, perhaps just for a motivating ground---can indeed be called a practical rule, but never a moral law.
>
> ---Immanuel Kant[2]

The above quotes highlight a sharp methodological difference between two philosophical approaches to morality. While Kant argues that moral obligations can only be grounded in *a priori* justifications rather than in anything empirical, Joseph Butler aims to ground moral obligations in empirical knowledge about human beings. Since they cannot both be correct, in this thesis, I aim to understand and evaluate their methodological conflict. Despite the apparent radical difference, I argue that Kant actually agree with Butler that moral obligations must be grounded in the understanding of human beings. Kant, however, fundamentally disagrees with Butler about human nature, which generates their methodological conflict in studying morality. For Kant, the most essential attribute for human beings is freedom, which is the capacity to set

1. Joseph Butler, *Fifteen Sermons and Other Writings on Ethics*, ed. by David McNaughton (Oxford University Press, 2017), P, 12 & 13. All quotations from Butler are from David McNaughton's edition of *Fifteen Sermons and Other Writings on Ethics* (Oxford, 2017). References to Butler's sermons and their preface are given in the following manner respectively: S2, 10; P, 12. The part before the comma refers to the number of the sermon or the preface; the part after the comma refers to the numbering of the paragraph.

2. Immanuel Kant, *The Groundwork of Metaphysis*, 4:389. All citations of Kant's works will be based on the Academy edition.

one's own ends without the influence of any particular experience, i.e inclinations, desires, feelings. In contrast, Butler understands human nature as a system that includes different particular experiences. Although there is no conclusive answer about the correct understanding of human nature, I suggest that Butler's account of moral obligation is a plausible one that can be considered as a counterexample to Kant's strong position that only a priori justification can ground moral obligations.

## 2 KANT'S REJECTION OF ALL EMPRICAL APPROACHES

### 2.1 Ground moral obligations on absolute necessity

Kant rejects all empirical propositions as the justificatory ground for moral obligations because he believes that moral obligations must have "absolute necessity" (G, 4:389) and "it is an outright contradiction to want to extract necessity from an empirical proposition" (CPracR, 5:12).[3] Two questions arise with Kant's rejection. First, what does Kant mean by "necessity?" Second, how plausible is Kant's claim that moral obligations must have the kind of necessity he has in mind? Since the answer to the second question depends on the answer to the first question, I will start with articulating Kant's conception of necessity in his moral theory. I argue that Kant holds a very strong conception of necessity that does rule out all empirical justifications.

In the preface of the *Groundwork,* Kant says:

> Everyone must admit that a law, if it is to hold morally, must carry with it *absolute necessity* [emphasis added]; that the command: thou shalt not lie, does not just hold for

3. Kant does not have to commit to the thesis that *identification* of the supreme principle of morality must be purely a priori. In the *Groundwork I*, Kant seems to identify the supreme principle from our common sense of a good will. What Kant clearly holds is the thesis that the *justification* of the supreme principle of morality must be purely a priori. This paper concerns solely on the second thesis. For further discussion on this point, see Paul Guyer, *The Virtues of Freedom: Selected Essays on Kant* (Oxford University Press, 2016). Philip Kitcher. "'A Priori'." In *The Cambridge Companion to Kant and Modern Philosophy*, edited by Paul Guyer. Cambridge University Press (2006).

> *human beings only*, as if *other rational beings* [emphasis added] did not have to heed it; and so with all remaining actual moral laws; hence that the ground of the obligation here must not be sought in the nature of the human being, or in the circumstances of the world in which he is places, but a priori solely in concepts of pure reason, and that any other prescription that is in some certain respect universal, in so far as it relies in the least part on empirical grounds, perhaps just for a motivating ground---can indeed be called a practical rule, but never a moral law.[4] (G, 4:389)

The first sentence in the passage seems to imply that Kant holds the absolute necessity of moral obligation as an assumption with which everyone agrees. I will come back to discuss whether this assumption is as obvious as Kant thinks, but the central issue for now is to figure out what Kant means by necessity. I will simplify the task by not focusing on providing a positive account of Kant's conception of necessity. Instead, I tackle this issue from the opposite end by asking: What counts as contingent for Kant in his moral theory? There are two advantages for this approach. On the one hand, we do not need to engage the daunting task of articulating Kant's full-blown view of necessity. On the other hand, we can better pinpoint Kant's argument by transferring the negative claims such as "hypothetical imperatives do not have absolute necessity" into a positive one "hypothetical imperatives are always contingent." Then we can ask: In what sense are they contingent? Why does this contingency rule out moral obligations?

Contingency of a moral rule comes in different senses. A moral rule is contingent if it constrains me in *some situations* but not in other situations; if it constrains *me* in all situations but not others; if it constrains *actual rational beings*, i.e. human beings, but not counterfactual rational beings in some possible worlds. As the previous passage shows, Kant believes that morality is contingent if it just holds for human beings but not for other rational beings. This means, as I shall argue, any *particular experience*, i.e. feelings, desires, inclinations, is

4. For similar discussions of rational beings as such in the *Groundwork*, see 4:408, 4:411-3, and 4:425-6.

contingent to rational beings, who are the appropriate subjects of morality. I must clarify a potential misunderstanding right away. I do not claim that *having experience* is contingent to rational beings. This is clearly false according to Kant, who always acknowledges the finite nature of rational beings. To illustrate this point, let me start with the concept of *pure rational beings.* Kant hardly ever uses this term, but I introduce it to avoid some unnecessary semantic issues. A pure rational being is someone who only has the faculty of pure reason. In contrast, a *rational being* is someone who has not only the faculty of pure reason, but also other sensible faculties. Even a *holy being* does not have to be a pure rational being. A holy being might be someone whose inclinations and pure reason always line up in determining a will. For a holy being, there are no imperatives and obligations because imperatives and obligations presuppose the conflict between subjective inclinations and objective laws recognized by pure reason (G, 4:414). A human being, who belongs to the subset of rational beings, is someone whose subjective inclinations often conflict with the objective laws. Since the conflict is required, Kant's moral theory presupposes the necessity for rational beings to have inclinations, desires, interests and other sensible experience.

However, the necessity of having experience does not imply that having any *particular* inclinations, desires or interests is necessary for a rational being. For example, as human beings, we find pleasure desirable and suffering undesirable. But a rational being, with a different cognitive and conative structure, might find pleasure undesirable and suffering desirable. In other words, some rational beings might have totally different inclinations as human beings normally have. Therefore, any moral rules derived from particular inclinations will be contingent because all particular inclinations are contingent. The earlier quote from Kant hints his belief that particular inclinations are contingent. Kant claims that a rule holds universally for human

beings is not automatically qualified as a moral law because it might not hold universally for all rational beings. This claim presupposes the contingent nature of human beings. Otherwise, commands that hold universally for human beings would necessarily hold for rational beings.

Kant's other works further demonstrate his belief that any particular inclinations are contingent. In his lectures, Kant says:

> But morality simply does not admit of being felt. All rules derived from feeling are contingent, and valid only for beings that have such a feeling. Feeling is a satisfaction that rests on the constitution of a sense. So it would then be all one, if God had also framed in us a liking for vice, and then He might equally have done it in other creatures as well. Such laws are therefore merely arbitrary, and simply a childish game. (LE, 29:625)

In this passage, Kant implies that all particular feelings are contingent for rational beings because feelings rest on having a specific way of sensing the world. For Kant, since the senses we use to generate feelings can be constituted in radically different ways for other rational beings, and a universal law must also hold for other rational beings, no moral laws can be derived from particular feelings. In other words, if we assume a feeling can be necessary for rational beings, then the rules derived from this feeling should also be necessary. But Kant clearly rejects this. A similar idea can be found in the *Critique of Practical Reason*. Arguing that a law must be free from any material or empirical condition, Kant states that particular experience such as a sympathetic sensibility cannot be presupposed in all rational beings because *God* does not necessarily have it (CPracR, 5:34). Kant suggests a very strong test condition for things that are necessary for rational beings. If we cannot presuppose a particular feeling in God, then the feeling is contingent thus is not appropriate for establishing a law. Kant's exact conception of God is not important here, but it is fair to say that God is very close to a pure rational being. Given this conception of contingency, Kant can rule out all empirical approaches

in grounding moral obligations because empirical approaches rely on identifying particular experience that is contingent for rational beings as such.

Let me address a potential objection to my interpretation, which seems to be implied by Kant's discussion of happiness. In the *Groundwork*, Kant seems to treat happiness as necessary for rational beings:

> [T]here is one end that can be presupposed as actual in all rational beings, and thus one purpose that they not merely can have, but that one can safely presuppose they one and all actually do have according to a *natural necessity*, and that is the purpose of happiness……One must present it [happiness] as necessary not merely to some uncertain, merely possible purpose, *but to be a purpose that one can presuppose safely and a priori in every human being, because it belongs to his essence*…. Thus, the imperative that refers to the choice of means to one's own happiness, i.e. the prescription of prudence, is still hypothetical; the actions is not command per se but just as means to another purpose [my emphases]. (G, 4:416)

In this passage, Kant claims that happiness is not contingent for human beings, which opens the possibility that the inclination for happiness is not contingent. If the inclination for happiness is necessary, then the principle of happiness, which is a hypothetical imperative---you should do X because it promotes your happiness---seems to have the equal binding force as the categorical imperative---you should do X because it is your duty. The prima facie challenge for Kant is to demonstrate exactly why this hypothetical imperative of prudence cannot bind us to the same degree as categorical imperatives bind us.

One might suggest the following reply on behalf of Kant: the imperative from the principle of happiness is hypothetical, and there is a fundamental distinction between hypothetical and categorical imperatives. Since the former is conditional and the latter is not, the former can only counsel while the latter can command. This intuition, however, is not self-evident for me. In fact, as hinted earlier, I have the opposite intuition. If the inclination for happiness is absolutely necessary, then the hypothetical imperative of prudence should bind as

equally as the categorical imperative does. I do not think Kant relies simply on this intuition. Instead, he distinguishes the two kinds of imperatives according to the degree of contingency and rules out the principle of happiness based on the strong conception of contingency I previously discussed. Right after the above passage, Kant states:

> [O]nly the law carries with it the concept of an unconditional and indeed objective and hence universally valid necessity, and commands are laws that must be obeyed, i.e. must be complied with even contrary to inclination. *Giving counsel does indeed contain necessity*, but it can hold only *under a subjective contingent condition*, if this or that human being counts this or that as belonging to his happiness; whereas the categorical imperative is limited by no condition, and *as absolutely and yet practically necessary* can quite properly be called a command [my emphases]. (ibid.)

As Kant acknowledges, the hypothetical imperative of happiness, which can counsel but not command, does entail a sense of necessity. What it lacks is the absolute necessity that morality requires. This imperative lacks the absolute necessity because it relies on particular desire, which is contingent. Although happiness as an abstract concept is necessary for all rational beings, it nonetheless depends on contingent sensible faculty because each of us identifies happiness with the "particular feeling of pleasure and displeasure" (CPracR, 5:25). Therefore, the principle of happiness cannot be a command because it relies on particular feeling, which is always contingent for rational beings based on Kant's strong conception of contingency. Thus, the necessity of happiness does not contradict Kant's claim that any particular experience is contingent for rational beings as such.

### 2.2 Ground absolute necessity on autonomy

If we agree with Kant's conception of contingency and his claim that morality must have absolute necessity, then we should agree with him that no empirical approach is capable of grounding moral obligation. To evaluate the plausibility of Kant's claim, we need to first find out his arguments for the claim. Kant, however, does not provide explicit arguments. He seems

to hold this claim as an uncontroversial assumption. But given Kant's strong conception of contingency, we cannot accept it too quickly. This situation forces us to reconstruct an argument for Kant's assertion. In what follows, I will draw on Paul Guyer's interpretation of Kant to suggest a potential argument for Kant's rejection of all empirical approaches based on the necessity of morality. I choose Guyer's interpretation for two reasons. First, I think his interpretation is very convincing and probably correct. Second, Guyer's interpretation will provide us a common ground to compare the methodological difference between Kant and Butler. Two claims in Guyer's interpretations are especially relevant to our current discussion. First, Kant derives "ought" from "is." Put differently: Kant establishes moral obligations from descriptive facts of rational beings. Second, Kant relies on the fact that freedom, or autonomy, is the most essential attribute for rational beings. If we do not have autonomy, we are no more different than animals.

It is commonly thought that Kant follows David Hume's sharp distinction between "ought" and "is," accepting that no normative principles can be derived from descriptive statements because they have entirely different foundations. Guyer, however, argues that Hume and Kant do not hold this view. What they hold is that "ought" must be derived from the right kind of "is." Normative principles cannot be derived from the objective facts about the external world, like what rationalists try to do. Instead, the normative principles can be derived from facts about human beings (2016, pp. 21-35). We normally approach descriptive facts of human nature from an empirical perspective, trying to understand the psychological and biological aspects of human beings. Kant, on the other hand, stresses the inadequacy of such empirical approaches. For him, some aspects of human beings can only be understood a priori. Kant distinguishes two kinds of descriptive facts about human nature. I will use *rational nature* to refer to the

description that cannot be understood with empirical approaches. It is a description of our noumenal self. I will use *non-rational nature* to refer to what Kant usually calls human nature, or the descriptions about the phenomenal self. Kant rejects all moral theories that are based on our non-rational nature because these theories ignore the most essential feature of human nature---autonomy---that can only be found in the noumenal self.

For Kant, the rational nature that distinguishes rational beings from animals is the capacity to set one's end. In the *Groundwork*, he explicitly says "a rational nature is distinguished from the others by this, that it sets itself an end" (G, 4:437). In the *Metaphysics of Morals*, he illustrates the same point with a slightly different term---humanity. As Kant points out, "the capacity to set oneself an end---any end whatsoever---is what characterizes humanity (as distinguished from animality)" (MM, 6:392).[5] This capacity to set one's end presupposes freedom, both in its negative and positive senses. The negative freedom allows rational beings to will without the influence of passions, inclinations and feelings. These sensible dispositions are "the monsters one has to fight" (MM, 6:405). The positive freedom, which is the autonomy, allows the will to be a law to itself (G, 4:440). In other words, if a will is autonomous, its sole determination is the will itself. The autonomous will is in a sense the unmoved mover. The autonomous nature of rational beings is controversial due to the seemingly logic contradiction of unmoved mover, I will come back to this point later. The upshot of previous discussion is that for Kant, we are not rational beings without freedom.

Even if rational beings are autonomous, why must morality be absolute necessary? I will answer this question by asserting a famous and widely accepted thesis coined by Henry Allison---the Reciprocity Thesis. According to this thesis, a free will and a will under moral laws are

5. Earlier in the *Metaphysics of Morals,* Kant writes "…More and more toward humanity, by which he alone is capable of setting himself ends" (MM, 6:387).

identical. As Kant states in the third section of the *Groundwork*, "if freedom of the will is presupposed, morality along with its principle follows from it, by mere analysis of its concept" (G, 4:447). Being rational beings requires freedom, which presupposes the exclusion of any particular experience. Since being rational is equivalent to being under moral laws, morality also requires freedom. Therefore, morality itself must be necessary in the sense that any particular experience is contingent. In other words, since any contingent element will deprive us from being free, which is identical to being under moral laws, all empirical approaches are not appropriate for grounding moral laws. One difficulty for Kant's descriptive account of rational nature is his metaphysical conception of autonomy. The transcendental nature of autonomy has generated controversy for centuries. Since Kant's rejection of all empirical approaches depends on his account of rational nature, an evaluation of his methodological commitment relies on the evaluation of his descriptive account of human nature. I shall postpone the evaluation until I lay out a competing account of human nature from Joseph Butler. Having both accounts helps us see the difficulty of evaluating which account is better.

## 3 BUTLER AS A POTENTIAL COUNTEREXAMPLE TO KANT

If my analysis so far is correct, then we see that Kant's insistence on a pure a priori approach to morality's normative foundation is rooted in his belief that human beings are autonomous. Joseph Butler shares the same fundamental assumption with Kant---moral obligations must be grounded in the descriptive facts about human nature. They disagree about the basic account of human nature. While Kant takes the essential part of human nature to be autonomy, which enables the will to be determined by itself without any influence from sensible experience such as inclinations, emotions and feelings, Butler treats human nature as a system. Distinguishing passions, self-love, benevolence and conscience as different parts of our nature and articulating their relations, Butler famously argues that conscience has supreme authority over other parts, so we ought to act in accordance with conscience. For Butler, to understand the system is to understand the relations between different parts, which necessarily include the non-rational nature. Therefore, an empirical approach is necessary to justify moral obligations according to the method chosen by Butler.

Although most commentators agree that the doctrine of the supremacy of conscience is the most distinctive element of Butler's moral philosophy, they disagree about what arguments Butler provides to defend this doctrine and whether his arguments are convincing. A central issue for using Butler's account as a counterexample to Kant's account is whether Butler successfully grounds moral obligations in non-rational human nature. In order to be counted as a counterexample, Butler's account needs to accomplish two things. First, it needs to establish normativity. Second, it must derive moral obligations from the non-rational nature of human beings. If Butler grounds moral obligations with a disguised rationalist account based on abstract facts about the world, then his account is not a counterexample. However, this does mean

Butler's account must be free of a priori propositions. Since Kant holds a strong view that nothing empirical can be used to justify moral obligation, as long as something empirical is necessary for Butler's account, this account will be a potential counterexample to Kant.

Butler does achieve the two things mentioned above. I shall argue that Butler's arguments for the supremacy of conscience derive directly from conscience's non-normative functional role to regulate other parts of our nature. Although it is debatable whether normativity can be derived from descriptive facts at all, Butler does not rely on any assumption that is less plausible than that of Kant. I will postpone the discussion of normativity to the next section, the main task for this section is to restore Butler's account of human nature and his attempt to derive moral obligations from it. The unique feature of Butler's moral theory is that it does not rely on any intrinsic value of a particular faculty. Although he argues that conscience is the supreme feature of human nature, he arrives at this conclusion by understanding the relationship between conscience and other parts of our nature. I believe previous interpretations have overlooked this perspective of Butler's account. My interpretation of Butler concentrates on his emphasis of human nature as a system and the relationship between different parts.

### 3.1 Non-relational function versus relational function

I shall argue that the supremacy of conscience is grounded in its regulative function, which is relational. Let me start with some examples. The most familiar example for contemporary readers is a computer. Every computer has a central processing unit (CPU), which constantly collects information sent from other parts such as the keyboard, the mouse, and the graphics cards. The CPU also sends instructions back to those parts after performing basic arithmetic, logical and input/output operations. In this sense, the CPU regulates other parts of the computer. Obviously, Butler never saw a computer, so let us look at some examples around the 18th

century. A mechanical watch is the example Butler uses. He claims that we can never understand the idea of a watch by focusing solely on its parts. We can only understand the idea when we "consider the respects and relations" the different parts "have to each other" (P, 14). Although it is unclear whether Butler considers any part of a watch to be regulative, regulation is important for a watch to function properly. The escapement wheel is the part that regulates the energy input from the mainspring. It makes sure the energy from the mainspring is released at a steady pace. Another 18th-century example is a centrifugal governor.[6] The crucial function of a governor is to regulate the speed of a mill or other energy source, and transferring the energy into controlled motions. In a drawing of centrifugal governor, James Watt titled it as "Centrifugal speed regulator."[7] This example shows the close connection between the regulative function of a part and its special importance as a governor of the entire system. In a broad sense, for Butler, conscience is similar to the CPU, the escapement wheel and the centrifugal governor. It regulates other parts of our nature, including self-love and benevolence.

These examples also highlight a distinction between my approach and previous commentators' approach. I emphasize conscience's relational function rather than its non-relational function. In the previous example, the CPU's regulative function is relational because it cannot collect data from other parts and send instructions back to those parts without interacting with them. In contrast, the CPU's calculating operation is non-relational because this function does not require interaction with other parts. Once the data are collected, the CPU performs calculation by itself. According to the non-relational approach, in a system, part A is superior to part B because A has non-relational function X and B either lacks X or has non-

6. Thank Dr. Eric Wilson for the example.

7. See S. Bennett (1993), *The History of Control Engineering 1800-1930*, pp.13.

relational function Y that is inferior to X. In order words, we establish the superiority between A and B based on the superiority between their non-relational functions X and Y. The non-relational approach is the default interpretation of Butler's argument for the supremacy of conscience. Commentators offer different views of which non-relational function accentuates conscience's superiority. One common interpretation focuses on conscience's ability to "pronounce determinately some actions to be in themselves just, right, good; others to be in themselves evil, wrong, unjust" (S2, 8). Other parts such as self-love clearly lack this function of making moral judgements.[8]

Self-love, however, has its own unique function—to secure one's own happiness with the best means. The unsettled problem for the non-relational approach is to show that the function of conscience is superior to the function of self-love. There seems to be no other way but relying on the overridingness of moral judgment. Terence Penelhum follows this line of thought. He argues that the function of making moral judgments is superior because "each of us is the sort of creature who is endowed with a recognition of the fact that the judgment of rightness or wrongness embodies an overriding reason, and that we should disregard the resistance to obligation of which the question is a sophisticated expression" (1985, p. 71). The overridingness of moral judgment, Penelhum continues, is a kind of knowledge or intuition that Butler believes

8. A non-relational approach might focus on the reflective power of conscience. P.F. Brownsey (1995, pp. 71-2) correctly notes the problem for this approach. Although it could be argued that self-love is superior than passions because self-love is reflective while passions are not, it is unclear how to establish conscience's superiority over self-love since both of them are reflective faculties. Terence Irwin (2008, p. 513) argues that conscience expresses the reflective function more *fully* than self-love. However, Irwin's account ultimately relies on conscience's ability to reflect upon "my nature as a whole," which self-love lacks (ibid., p. 524). Reflecting upon nature as a whole presupposes the relational regulative function between conscience and other parts of our nature.

everyone has. So "the onus [of proof] is on those who think some other clear sense exists, to elucidate it, and that their success is not conspicuous" (p. 73).

Penelhum's interpretation has at least two problems. First, it reduces the normative force of Butler's account to a Clarkean rationalist approach, which obscures the distinctness of Butler's approach. Based on this interpretation, the sole function of conscience is to grasp the abstract moral facts in the world, which grounds the overridingness of moral judgment. Although Butler claims that his account and Clark's account "both lead us to the same thing" (P, 12), he does not claim that his account depends on Clark's account. Butler only says that their approaches are compatible. Immediately after introducing conscience's function of making moral judgments, Butler states "But this part of the office of conscience is beyond my present design explicitly to consider" (S2, 8). The "present design" for Butler is to establish the normative authority of conscience. And he wants to alert us here that he is not trying to achieve this goal by relying on conscience's ability to make moral judgments. Elsewhere, Butler states:

> Yet let any plain honest man, before he engages in any course of action, ask himself, 'Is this I am going about right, or is it wrong? Is it good, or is it evil?' I do not in the least doubt but that *this question would be answered agreeably to truth and virtue* [emphasis added], by almost any fair man in almost any circumstance." (S3, 4)

Butler notes that virtue and moral judgments coincide, which implies that virtue is not derived from the judgments of right and wrong, for if right moral judgments *define* virtue, it will be redundant to point out that moral judgments and virtue always agree. Second, Penelhum's interpretation imposes a dilemma on Butler. According to his interpretation, the normative force of conscience depends on its ability to make correct moral judgments. The moral judgments of conscience are either fallible or infallible. If they are infallible, how can Butler defend this awfully strong claim? If they are fallible and we can know when they are false, as Penelhum holds, then there must be another faculty distinct from conscience that can discern the mistakes

made by conscience. If this "other" faculty exists, it is unclear why Butler holds conscience as the supreme faculty instead of the "other" faculty.

These problems of the non-relational approach plus Butler's commitment to understand human nature as a system motivate my relational approach. The relational approach establishes superiority based on the relation part A and part B bear. A is superior to B just because A bears a regulative relation, which is asymmetric, with B. For example, the CPU is superior to other parts of a computer such as a mouse or graphics cards not because of its unique function of performing basic arithmetic operations. The CPU is superior to other parts because it constantly monitors and collects data from *other parts* and it sends instructions *back to these parts*, and *not vice versa*. The CPU cannot accomplish the regulative function without interacting with other parts.[9] As far as I know, no one has fully developed this interpretation of Butler. The closest account is suggested by Ralph Wedgwood. He draws our attention to a crucial passage where Butler states "you cannot form a notion of this faculty, conscience, without taking in judgment, direction, superintendency. This is a constituent part of the idea, that is, of the faculty itself" (S2, 14). I will quote Wedgewood's insightful interpretation of this passage in length:

> This argument seems to start from identifying certain features of conscience. One feature is conscience's "superintendency," which seems to consist in the way in which conscience considers and oversees all the other internal principles of the mind. Another feature is its capacity for "judgment", which may consist in the fact that it arrives at all-things-considered judgments about what is right or wrong, good or bad, aiming to take all relevant considerations into account. The final feature is its capacity for "direction", which may consist in the fact that we have a fundamental disposition to be moved to action by the directions of our conscience. (2007, p. 194)

9. The "superiority" of the CPU is not normative because it simply describes the non-normative asymmetric relational function of the CPU on other parts. One might try to establish the normative "superiority" from the non-normative functional relation, but we should not automatically equate the two senses of superiority. I focus solely on the non-normative relation in this section.

I agree with Wedgewood's interpretation of "superintendency" and "direction." Overseeing other internal principles underscores the relational function of conscience over other faculties. However, I think Wedgewood misinterprets what Butler means by "judgement" here. Butler is not talking about the non-relational function of making moral judgments, but the relational function of evaluating other principles. Judgment, direction and superintendency constitute conscience's regulative function. Because Wedgewood interprets "judgment" in an indirect way, he concludes that "it is not completely clear why Butler says that a faculty that has these features 'claims superiority from its very nature' (ibid.)" In what follows, I shall elucidate this regulative function in a finer grain, which helps us better understand Butler's account of human nature and his argument for the normative supremacy of conscience.

### 3.2 Self-love's regulative function

Butler starts the discussion of regulative function with self-love. The function of self-love is to regulate passions in order to secure happiness for ourselves. Butler realizes that it is not easy to distinguish different parts of our nature, so he believes "it will be necessary to consider the nature, the object and end of" different faculties (S11, 4). For Butler, the objects of passions are external things like food or esteem, whereas the object of self-love is one's own happiness. However, self-love is not a blind desire for happiness because it also "puts us upon obtaining and securing" happiness, which requires a reflective power to regulate passions (S11, 9).

To see why it is necessary for self-love to have regulative power, we need to know more about Butler's concept of happiness. For Butler, happiness "does not consist in self-love," it consists in "the gratification of particular passions" (S11, 9). In other words, if I am hungry and I get a good meal, I will be happy. Similarly, if I desire esteem and get into a prestigious PhD program, I will be happy. However, being happy is more complex than fulfilling an immediate

passion. First, although the gratification of particular passions always brings me immediate happiness, it might not bring me long-term happiness. Butler repeatedly emphasizes that a passion might motivate us to "rush upon certain ruin for the gratification of a *present* [emphasis added] desire" (S1, footnotes 3; S2, 10). Thus, in order to secure long-term happiness, self-love must make judgments when a person is about to act according to a passion. In the situation given by Butler, self-love will always judge against running into self-harm. In addition, we often have multiple passions that are potentially conflicting. In these situations, self-love's regulative function is necessary in order to help one achieve long-term happiness. As David McNaughton helpfully articulates:

> [T]he gratification of any affections gives us pleasure, but gratification of one often impedes the gratification of others. Our overall long-term happiness consists in harmonizing their gratification, so that each affection attains its object within its due stint and bound. This harmonizing is the task of self-love." (2017, p. xvi)

Thus, self-love is the faculty that superintends and judges upon different passions. Can self-love provide a judgment without overseeing and evaluating passions? No. Butler tells us that "take away these affections, and you leave self-love absolutely nothing at all to employ itself about" (P, 37).

### 3.3 Conscience's regulative function

Conscience's regulative function is essentially the same as self-love's regulative function, but Butler does not tell us explicitly what parts of our nature conscience regulates. I shall argue that the objects of conscience, for Butler, are self-love and benevolence. Butler first introduces conscience as:

> There is a principle of reflection in men, by which they distinguish between, approve and disapprove their own actions. We are plainly constituted such sort of creatures as to reflect upon our own nature. The mind can take a view of what passes within itself, its propositions, aversions, passions, affections, as respecting such objects, and in such degrees…This principle in man, by which he approves or disapproves his heart, temper,

> and actions, is conscience; for this is the strict sense of the word, though sometimes it is used so as to take in more. (S1, 8)

This passage can be read in two ways. One can stress conscience's function of approving and disapproving *actions* based on right and wrong. This will lead to the non-relational approach. I have argued that Butler does not take this approach. Instead I think Butler emphasizes the relational function of conscience, which "reflects upon our own nature." Butler almost always describes conscience as a faculty that judges other parts of our nature. Conscience distinguishes between the "internal principles of his heart" (S2, 8). Conscience "approves or disapproves the several affections of our mind" (S8, 9).

Although Butler does not explicitly say that conscience regulates self-love, he is likely to hold this view. To illustrate, we need to note two things about self-love. First, self-love is just one part of our whole nature and it has flaws. Second, self-love can be in conflict with benevolence. Butler states, "How much soever therefore is to be allowed for self-love, yet it cannot be allowed to be the whole of our inward constitution; because, you see, there are other parts or principles which come into it" (S11, 8). The distinction between nature and whole nature is important. For Butler, the most basic sense of "nature" is just "some principle in man, without regard either to the kind or degree of it" (S2, 5).[10] Thus, acting in accordance with passions, self-love, benevolence and conscience are equally natural in this basic sense. However, when he claims that virtue is following our nature, he is not using "nature" in the basic sense. He is talking about our whole nature, which includes conscience's regulation upon self-love. Moreover, Butler warns us about the self-defeating feature of self-love, which is made known by Henry Sidgwick as the "fundamental paradox of egoistic hedonism." The more we focus on our

10. "Principles" for Butler should be understood as passions, inclinations or feelings, which are very different from Kant's use of the term. Principles for Kant are rules or laws used for deriving maxims.

own happiness, the less happy we will be. So "disengagement is absolutely necessary to enjoyment" (S11, 9). How can we disengage from the obsession of self-love? Since self-love cannot oversee itself, conscience needs to oversee self-love, making sure we are not obsessed with it.

Self-love has a counterpart. "Benevolence," Butler states, "is in some degree to society, what self-love is to the individual" (S1, 6). While self-love is a reflective faculty that puts us upon securing happiness for ourselves, benevolence is a reflective faculty that puts us upon securing other's happiness. It is not hard to imagine situations where self-love and benevolence conflict. For example, while on her way to an important meeting, Bobo sees someone in need of help. If Bobo helps him, she would not be able to make the meeting, which would have considerable negative effects on her career. In situations like this, since self-love reflects solely on Bobo's happiness and benevolence reflects solely on the happiness of the person who needs help, Bobo cannot make a harmonious decision without relying on conscience's judgment upon self-love and benevolence.

Butler, however, seems to believe that self-love and benevolence rarely contradict each other. If this is true, it will undermine my interpretation because there is no need for conscience to regulate. To answer this objection, we need to pay attention to Butler's distinction between benevolence as a passion and benevolence as a reflective principle:[11]

> Love of our neighbor is one of those affections. This, considered as a virtuous principle, is gratified by a consciousness of endeavoring to promote the good of others; but considered as a natural affection, its gratification consists in the actual accomplishment of this endeavor. (S11, 16)

11. For a similar discussion, see C.D.Broad (1959), *Butler,* pp. 71-73 and Terrence Irwin (2008), *Butler,* pp. 507-510.

Benevolence as a passion is just a desire to promote other's happiness, which is no more different than passions for food or esteem. Similar to other passions, benevolence is also an "instrument of private enjoyment" because the gratification of it will bring one happiness. Thus, Butler holds that "there is no peculiar rivalship or competition between self-love and benevolence" (S11, 19).

An obvious weakness of Butler's argument is that it exaggerates our passion of benevolence. Most people are not as benevolent as Butler thinks, so it is unclear how much happiness they can gain by promoting others' happiness. However, for the sake of argument, let us accept Butler's assertion. Nonetheless, I agree with Terrence Irwin (2008, p. 524) that benevolence as a reflective principle might conflict with self-love. The example of Bobo offers a situation of such conflict. Furthermore, Butler's discussion on how much we should love our neighbors underscores the potential tension between benevolence and self-love. For Butler, one must love her neighbors and herself in the same degree (S12, 6). Given this strong requirement of benevolence, it is not hard to think about situation where one cannot promote other's happiness and one's own happiness simultaneously.[12] Therefore, conscience, for Butler, is the faculty that regulates self-love and benevolence in order to secure the harmony between promoting one's own and other's happiness.

Even if I am correct that conscience oversees and judges self-love and benevolence, one might object that conscience is not superior to self-love because self-love also regulates conscience. Scholars often cite the following passage to emphasize the supreme role of self-love:

12. In addition, as Wedgewood (2007, pp.203-4) convincingly argues, the harmony of self-love, benevolence and conscience is the least compelling claim for Butler's account. So, if we have to stretch Butler's account for greater plausibility, it is best to reject this claim and highlight the potential conflict between different parts of our nature, which speaks in favor of his more important doctrine, that is, the supremacy of conscience.

> And to all these things may be added, that religion, from whence arises our strongest obligation to benevolence, is so far from disowning the principle of self-love…It may be allowed, without any prejudice to the cause of virtue and revision, that our ideas of happiness and misery are of all our ideas the nearest and most important to us; that they will, nay, if you please, that they ought to prevail over those of order and beauty, and harmony, and proportion, if there should ever be, as it is impossible there ever should be, any inconsistence between them…Let it be allowed, though virtue or moral rectitude does indeed consist in affection to and pursuit of what is right and good, as such: yet, that when we sit down in a cool hour, we can neither justify to ourselves this or any other pursuit, till we are convinced that it will be for our happiness, or at least not contrary to it. (S11, 20)

As Irwin (2008, p. 709) and others note, "the appeal to self-love is a concession ("let it be allowed") made for the sake of argument."[13] I agree with this interpretation. Butler does not state his view of the proper role of self-love here. He tries to argue that *even if* one holds self-love as the supreme principle, there is no practical harm because self-love and benevolence do not conflict. Nonetheless, the previous passage focuses on the normative role of self-love instead of its non-normative function. It does not undermine my interpretation because all I aim to show is that self-love does not regulate conscience in a non-normative sense.

One might argue that, however, when deliberating over conflicts of benevolence and self-love, we seem to experience that self-love reflects upon conscience. In the example of Bobo, after her conscience determines that she should help the strangers, she might reconsider this determination and reflect whether helping the stranger really contributes to her own happiness. Despite this intuition, I do not think self-love actually *regulates* conscience. On the one hand, self-love can only judge about one's own happiness, and it has already concluded that going to the meeting is better than helping the stranger. Conscience takes the judgment of self-love into consideration and evaluates self-love against benevolence. The final verdict of conscience generates no new information that deserves further reflection from self-love. For example,

13. Also see Sturgeon (1976), *Nature and Conscience*, p. 338; Broad (1959), *Butler,* p. 80; Wedgewood (2007), *Butler on Virtue*, p. 197.

announcing a verdict against one party, the Supreme Court already know that the verdict is against this party's interest. So, it is redundant that the party reflects again upon the verdict and claims "this decision is against my interest." On the other hand, what we feel as self-love's regulative function might just be self-love's motivational strength or its uncompromising self-affirmation. As Butler notes, "Reflection or conscience comes in, and disapproves the pursuit of them in these circumstances; *but the desire remains* [emphasis added]" (S2, 13). What we feel might just be the "remaining desire" of self-love, instead of self-love's judgment upon conscience's verdict.

### 3.4 Deriving normativity from regulative function

Given conscience's regulative function, why should we act in accordance with it? In other words, where does the normative authority of conscience come from? I submit that Butler derives the normative authority directly from the regulative function. He starts the discussion of the normative hierarchy of human nature with self-love:

> So that, if we will act conformably to the economy of man's nature, reasonable self-love must govern. *Thus, without particular consideration of conscience, we may have a clear conception of the superior nature of one inward principle to another* [my emphasis]; and see that there is this natural superiority, quite distinct from degrees of strength and prevalence. (S2, 11)

The superiority of self-love to passions indicated in the passage is clearly normative and Butler claims that we can establish the superiority without relying on conscience. Butler's claim reinforces my interpretation that the normative superiority does not depend on the function of making moral judgments because self-love does not even have this function. The normativity is derived from self-love's regulative function over passions. Butler emphasizes that natural superiority obtains in the relation of "one inward principle to another" (S2, 11-12). The relational approach fits well with this description. Based on the relational approach, superiority only makes

sense in a comparative context. Self-love is superior to passions because self-love regulates passions. "Self-love is superior" by itself does not express anything substantive.

Similarly, conscience is superior to self-love and benevolence because it regulates them. Butler moves from self-love's superiority to conscience's superiority rather quickly because he thinks the regulative functions of conscience and self-love are parallel. And once we understand the superiority of self-love, we understand the superiority of conscience. This brings us back to the crucial passage identified by Wedgewood:

> As from its very nature manifestly claiming superiority over all others: insomuch that you cannot form a notion of this faculty, conscience, without taking in judgment, direction, superintendency. This is a constituent part of the idea, that is, of the faculty itself: and, to preside and govern, from the very economy and constitution of man, belongs to it. (S2, 14).

As Wedgewood correctly points out, for Butler, conscience's normative authority comes from its abilities to judge, direct, and superintend. What Wedgewood finds unclear is Butler's jump from these abilities to conscience's normative supremacy. The relational approach sheds some light on understanding Butler's argument. What Wedgewood overlooks is the direct relation between conscience and other parts. Conscience not only judges, directs, and superintends, but does all these *upon self-love and benevolence*. Conscience is superior to self-love because self-love is the subject-matter of conscience and not *vice versa*. The asymmetrical regulative relation between conscience and self-love gives conscience the normative authority. The normative superiority for Butler is never unconditional as "A is supreme." It always has the form "A is *superior to* B." Thus, the regulative function of conscience does not immediately lead to the unconditional obligation---we should act in accordance with conscience. Instead, it leads to the conditional obligation---we should act in accordance with conscience rather than self-love. The unconditional obligation, or the supremacy of conscience, is established with a further fact that conscience is the faculty that judges "without being consulted and without being advised" (S2,

8). In other words, human nature as a whole, for Butler, does not have any other parts that are superior to conscience.

The discussion in this section is not an evaluation of the plausibility of Butler's normative theory. It simply presents Butler's intriguing account, which derives moral obligation from a description of human nature as a hierarchical system. This section aims to show that Butler's normative theory is empirical, which is prey to Kant's methodological rejection. Butler's approach is empirical because it relies on the empirical fact that human nature is hierarchical. In addition, since Butler's claim about human nature is universal, he relies on the generalization of particular examples of human behaviors, which is firmly rejects by Kant because it is an improper approach to ground moral obligations (G, 4: 408).

# 4 COMPARISONS BETWEEN KANT AND BUTLER

I have argued that both Kant and Butler aim to ground morality in the description of human nature. Their methodological difference on accepting empirical approaches in grounding moral obligations rests on their different understanding of human nature. Therefore, to evaluate whose approach is better, we need to address two issues: whose account is more successful in establishing normativity from description and whose account is a more accurate description of human nature. As I shall show in this section, neither of the two questions has an obvious answer.

I will start with the first issue. I think Kant and Butler face the same difficulty. If a description of human nature is not intrinsically normative, then as Guyer points out, we feel that an argument is needed to show that the natural is also normative (2016, p. 25). Even if Butler's hierarchical system and Kant's autonomy are inherent in human beings, one might argue that they are valueless or even dysfunctional. Thus, following our nature is not morally obligatory. It is either morally neutral or even morally objectionable. Guyer points out two ways to bridge this gap between description and normativity. The first way is to introduce a "teleological framework that will forestall any question about the normative force of the merely natural." He attributes this approach to Adam Smith. He argues that "Smith accepts an essentially Humean account of the origin of moral principles in sentiment rather than reason while situating our sentiments themselves in a teleological framework" (2016, p. 27). Once the teleological framework is introduced, no further explanation is needed to establish the normativity from the description.

Introducing a teleological framework is a common way to interpret the normative force of Butler's conscience. As Wedgewood argues, Butler holds that every system must have a proper function, which is determined by the way the system is "generally *disposed* to operate in"

(2007, p. 183). In other words, “a teleological conception of a system essentially incorporates a conception of what is the right or proper state for that system---that is, the state that the system ought to be in.” Based on this interpretation, once we find how the system is disposed to operate, we find its proper function of the system that ought to be followed. If we accept the teleological framework and my interpterion of Butler’s human nature, the normative hierarchy maps onto the descriptive hierarchy held by Butler, which states that self-love and benevolence regulate passions, conscience regulates self-love and benevolence, and no faculty regulates conscience. As a result, we justify the normative hierarchy, which states that we should act in accordance with self-love rather than passions; we should act in accordance with conscience rather than self-love; and conscience is the supreme principle we should follow.

However, for people who are skeptical about the bridge between description and normativity, why should they accept the teleological framework? For the skeptics, the teleological framework seems like an *ad hoc* fix of the problem rather than a genuine argument. The concern is that the teleological framework does not add anything valuable in term of the justificatory power. Even if it adds something valuable, a new concern is introduced because one might ask for the justification of accepting the teleological framework. If we keep providing arguments, there is a regress problem that can only be stopped by claiming something as self-evident. What argument counts as self-evident? I do not have a good answer, but I think my interpretation of Butler might suggest the possibility to derive the normality directly from the descriptive fact of human nature as a system, without relying on the teleological framework. A unique feature of the relational approach is that it entails a sense of order—self-love regulates passions and conscience regulates self-love and benevolence. Can this order be the foundation of moral obligation? If so, unlike the teleological approach which assumes the existence of a proper

end for human beings, Butler's account can start with a mere hypothesis of such an end. Through the exploration of human nature, we find the unique functional hierarchy, which is used for grounding moral obligations. Recall Butler's watch example. If we give a watch to a proficient technician who has never seen a watch, can she discover the function of the watch by studying its different parts and their relations? If she could, then we might discover our function by studying the different parts of human nature and their relations, which opens up a distinctive way to ground moral obligations that does not depend upon controversial theological or teleological assumptions.

Another way to bridge the gap between normativity and description suggested by Guyer is to introduce what Kames calls the "common nature." As Guyer points out, In *Elements of Criticism*, Kames claims that

> [T]he "common nature" is just that pattern to which the majority of the species happened to conform, and is not an intrinsically normative conception, but is nevertheless naturally "conceived" by us "to be a model or standard for each individual that belongs to the kind." (2016, p. 28)

In other words, the universality of a feature in human nature gives rise to the normativity. I doubt the universality can help us establish normativity. First of all, any investigation about human nature depends on finding something more or less universal. Kant's conception of autonomy and Butler's hierarchical system of human nature are both supposed to be universal for human beings. It is hard to accept a feature as our nature if the feature exists only for a small group of human beings. In other words, it is because some features are identified as "common nature," we have some confidence to treat these features as our human nature. The "common nature" can be used only for identifying our human nature, rather than providing support to derive normativity from human nature. Kames' suggestion puts the cart before the horse. Second, I think Kant correctly warns us that many aspects of our non-rational nature are contingent in a sense that we

should not let these aspects define who we are, especially if it is possible to change these aspects. If we look back into the history where human beings are more inclined to violence and discriminations against gender, race, and sexual orientations, we will not be confident to derive normativity solely from universal features of human beings. Learning from history, our current "common nature" might still be some outdated survival mechanisms for small tribal society. We need to make some kind of evaluative judgment about what the right descriptive facts are.

If the teleological framework and "common feature" do not provide valuable arguments for establishing normativity from descriptions, we might need to resist the skeptical thoughts and embrace the idea that we have to establish normativity from the right description of human nature. Nonetheless, there is still a significant difference between Kant's description and Butler's description of human nature. While Butler's description is essentially empirical, Kant's description is not. Is non-empirical description of human nature more suitable to ground moral obligations than an empirical description? This is the question Guyer recommends moral philosophers to focus on. He suggests further discussion in moral theory shifts its focus from "whether we should rigidly separate normative from descriptive approaches" to "whether within the latter we should prefer an empirical method that is heir to the method of Hume, Kames, and Smith, or whether we can employ an a priori method that is heir to the rationalist descriptivism of Kant's central period" (2016, p. 35). Since I am sympathetic to the idea that we can derive normativity only from some descriptive facts about who we are, I agree with Guyer's remark. In addition, I think the comparison between Kant and Butler's methodological difference provides an initial step to address the issue. If Kant is correct to point out that non-rational nature is more contingent than rational nature, then we might prefer Kant's description of human nature. Is Kant correct in holding this position? This question leads us to the discussion of the second

issue, that is, whether Kant provides a more accurate description of human nature than Butler does. A thorough comparison between Kant's and Butler's accounts of human nature exceeds the scope of this paper. I will make some brief remarks of the problems in Kant's and Butler's accounts, and the potential usefulness of empirical understanding of human beings.

Human nature is a highly contested concept. There is no consent about what the concept refers to. I will adopt some aspects of Maria Kronfeldner's framework to simplify the issue. Kronfeldner distinguishes two conceptions of human. *Humankind* is a biological group refers to *Homo sapiens*. This biological group is fixed by the genealogy, which includes intergenerational transmission of development resources and biological ancestor-descendent linages. *Humanity*, in contrast, is a social group whose membership is determined "by a set of traits that this group at a certain time regards self-referentially and discursively as necessary and sufficient for being a person and thus a member of humanity" (Kronfeldner 2018, pp.4-5). For Kant and Butler, the proper conception of human is *humanity* instead of *humankind* because not everyone in humankind is subjected to morality. The *human* nature that is important to them is the nature of humanity.

Another question is: what does it mean for a feature to be the *nature* of humanity? In other words, what are the traits or features that qualifies us to be the members of humanity, thus be the appropriate subjects of morality. As Leslie Stevenson points out, we should be alert to possible confusions when someone says "Human beings are naturally X." One might mean "All (or most) humans are X." or "Human beings ought to be X." or "Human beings have an innate tendency to X." or "Human beings were X before civilization."[14] My interpretation of Butler and Kant claims that they infer morality from human nature. Therefore, human nature cannot be the

14. Leslie Stevenson, *Thirteen Theories of Human Nature (*Oxford University Press, 2018*)*, p. 11.

normative conceptions of what human beings ought to be. The reason for not using a normative conception of human nature is to avoid a potential circularity. The central purpose for Kant and Butler is to justify moral obligations, which requires them to justify their claims such as we *ought to* act according to the law derived from the categorical imperative commands or we *ought to* act according to conscience. It is true that part of their projects is to explain why we ought to X instead of Y, but the crux is to explain where the normative force in the "ought to" comes from. If we infer the normative force of morality from the normative conceptions of human nature such as we ought to be autonomous beings or we ought to perfect the faculty of conscience, it seems to me that we only provide an explanation of the normativity in morality but not a justification. If we look for justification, using a normative conception of human nature traps us into a circle because we never justify "ought to" with something independent of "ought to." In order to avoid the circularity, the normativity of moral obligations must be justified with a descriptive conception of human nature, which refers to the universality or the typicality of humanity.

However, taking an empirical approach, we cannot identify the universal or typical features of humanity before identifying the category of humanity because we must know the subjects of the studies before studying them. This is the central difficulty for Butler. Thus, it seems that we can only understand the nature of humanity with the universal or the typical features of humankind, that is, H. sapiens. The problem for this approach is its temporal constraint because we can only study H.sapiens up to this time. Let's assume we have a full understanding of humankind up to this time, and we are able to identify the common nature shared by human beings as a species. Are we willing to call the shared features as non-contingent thus using them to define humanity? I do not think so because we have already

witnessed the change of the features shared by a large group of humankind for a relatively long period. Many of the universal features we have been possessing could be overwritten by other features we have not yet developed. These potential features can be equally desirable for humanity. Therefore, limiting the conception of human nature within the features of past humankind seems to be overly narrow.

Nonetheless, there is a potential way to strengthen Butler's and other empirical approach against the temporal constraints. As C. D Broad suggests, instead of discussing the actual nature, we can identify the ideal nature based on the actual nature. For example, we can form the conceptions of a perfect watch, perfectly straight lines, or perfect circles without such objects in Nature (1959, pp. 57-9). Although a somewhat mystical ability to abstract ideal nature is required, this ability is not counterintuitive. We always picture a better version of ourselves based on the past and current self. It seems that the more we know about ourselves, the easier it is to grasp the ideal version of ourselves. I think we can do the same thing for humanity. We reflect upon ourselves as data for ideal human nature. If this is true, then empirical understanding of human nature is essential because the process of idealization requires us to know actual human beings.

Kant's metaphysical approach seems to have some advantages because it does not have the temporal constraint. For Kant, rational beings are necessarily autonomous. However, can we confidently define human nature without any empirical understanding of us? Is autonomy the only essential part of us? Are we really willing to hold features such as conscience, love, care, or sympathy as merely contingent? I think we can actually hold Kant's emphasis on autonomy while accepting some empirical features as not contingent for humanity. The reason is that Kant's conception of autonomy does not rule out the possibility to *freely choose* conscience,

love, care and sympathy as the essential nature of humanity. Kant seems to think that choosing these features is not a "free choice" of pure reason because inclinations towards these features drive us to make the choice. Yet, if the defining feature of an autonomous being is the capacity to set her ends without being influenced by any particular experience, it must be possible to choose conscience, love, or sympathy without any inclination towards conscience, love or sympathy. If this is the case, we cannot rely solely on a pure a priori approach because we need empirically informed descriptions about us to help us make better choices. If it is correct that empirical knowledge of human nature plays a role in idealizing or choosing what constitutes humanity, then Butler's account of morality, which is derived from his description of human nature as a system, is a counterexample to Kant's position that moral obligations can only be grounded in *a priori* justifications.

## BIBLIOGRAPHY

Allison, H.E. (1986). Morality and freedom: Kant's reciprocity thesis. *The Philosophical Review*, 95(3), pp.393-425.

Akhtar, S. (2006). Restoring Joseph Butler's conscience. *British Journal for the History of Philosophy*, 14(4), pp. 581-600.

Bennett, S. (1993). *A history of control engineering 1800-1930*. London: The Institution of Engineering and Technology, p.13.

Broad, C.D. (1959). *Five Types of Ethical Theory*. New Jersey: Routledge.

Brownsey, P. F. (1995). Butler's argument for the natural authority of conscience. *British Journal for the History of Philosophy* 3(1), pp.57-87.

Butler, J. (1726). *Fifteen Sermons Preached at the Rolls Chapel and other writings on ethics*. In: McNaughton D. ed. *Fifteen Sermons and Other Writings on Ethics.* Oxford: Oxford University Press, 2017.

Darwall, S. (1995). Butler: conscience as self-authorizing. In: *The British Moralists and the Internal 'Ought': 1640-1740*. Cambridge: Cambridge University Press.

Guyer, P. (2006). Introduction to *The Cambridge Companion to Kant and Modern Philosophy*. Edited by Paul Guyer. Cambridge University Press.

Guyer, P. (2000). "The Strategy of Kant's Groundwork." In *Kant on freedom, law, and happiness*. Cambridge University Press.

Guyer, P. (2016). *The Virtues of Freedom: Selected Essays on Kant*. Oxford University Press.

Hill, T. (2002). "Kantian Analysis: From Duty to Autonomy." In *Human Welfare and Moral Worth: Kantian Perspectives*. Oxford University Press.

Kant, I. (2015). *Critique of Practical Reason*. Edited by Mary Gregor. Cambridge University Press.

Kant, I. (2012). *Groundwork of the Metaphysics of Morals*. Edited by Mary Gregor and Jens Timmermann. Cambridge University Press.

Kant, I. *Lectures on ethics.* (2001). Edited by Peter Heath and J.B. Schneewind. Translated by Peter Heath. Cambridge University Press.

Kant, I. (2017). *The Metaphysics of Morals*. Edited by Lara Denis. Translated by Mary Gregor. Cambridge University Press.

Kim, H. (2015). *Kant and the Foundations of Morality.* Lexington Books.

Kitcher, P. (2006). "'A Priori'." In *The Cambridge Companion to Kant and Modern Philosophy*, edited by Paul Guyer. Cambridge University Press.

Kronfeldner, M. (2018). *What's left of human nature? A post-essentialist, pluralist, and interactive account of a contested concept.* The MIT Press.

Irwin, T. (2008). Butler. In: *The Development of Ethics: Volume 2: From Suarez to Rousseau*. Vol. 2. Oxford: Oxford University Press.

MacIntyre, A. (1984). *After Virtue: A Study in Moral Theory*. Norte Dame: University of Notre Dame Press.

McNaughton, D. (2017). Introduction. In: Butler, J. *Fifteen Sermons and Other Writings on Ethics*. McNaughton, D. ed. Oxford: Oxford University Press, pp.xi-xxxvi.

Penelhum, T. (1985). *Butler: The Arguments of the Philosophers*. New York: Routledge.

Schneewind, J. (1992). "Autonomy, obligation, and virtue: An overview of Kant's moral philosophy." In *The Cambridge Companion to Kant,* edited by Paul Guyer. Cambridge University Press.

Stevenson, L. (2018). *Thirteen Theories of Human Nature.* Oxford University Press.

Sturgeon, N. (1976). Nature and Conscience in Butler's Ethics. *The Philosophical Review* 85(3), pp.316-356.

Wedgwood, R. (2007). Butler on Virtue, Self-Interest and Human Nature. In: Bloomfield. P. ed. *Morality and Self-Interest*. Oxford: Oxford University Press, pp.177-204.

Wood, A. (2006). "The supreme principle of morality." In *The Cambridge Companion to Kant and Modern Philosophy*, edited by Paul Guyer. Cambridge University Press.